AD-A056 827 STATE UNIV OF NEW YORK AT BUFFALO AMHERST STATISTICA--ETC F/G 12/1
NONPARAMETRIC STATISTICAL DATA MODELING.(U)
JAN 78 E PARZEN DAAG29-76-G-0239
UNCLASSIFIED TR-59 ARO-13845.6-M NL

1 OF 1
ADA
056827

END
DATE
FILMED
9 -78
DDC

State University of New York at Buffalo
Department of Computer Science

AD A056827

LEVEL II

ARO 13845.6-M

18 19

12

6 NONPARAMETRIC STATISTICAL DATA MODELING*

by

10 Emanuel Parzen

Statistical Science Division
State University of New York at Buffalo

An Invited Paper to be Published with Discussion in the
Journal of the American Statistical Association

12 67p.

9 GRANT TECHNICAL REPORT NO. ARO-3

STATISTICAL SCIENCE DIVISION REPORT NO. 59

11 January 1978 14 TR-59

15 DAAG29-76-G-0239

*Research supported by the Army Research Office (Grant DA AG29-76-G-0239).

409 511

78 07 21 026

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

A056827

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER ~~Technical Report No. ARO-3~~	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Nonparametric Statistical Data Modeling		5. TYPE OF REPORT & PERIOD COVERED Technical
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Emanuel Parzen		8. CONTRACT OR GRANT NUMBER(s) DA AG29-76-G-0239
9. PERFORMING ORGANIZATION NAME AND ADDRESS Statistical Science Division State University of New York at Buffalo Amherst, New York 14226		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS		12. REPORT DATE January 1978
		13. NUMBER OF PAGES 58
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

NA

18. SUPPLEMENTARY NOTES

The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Nonparametric Statistics
Exploratory Data Analysis
Autoregressive Density Estimation
Quantile Functions
Density-Quantile Functions
Spacings
Goodness of fit
Parameter Estimation
Tail of a Distribution

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

It is the aim of this paper to introduce new types of keys for exploratory data analysis (of continuous data) based on estimating the quantile function and density quantile function. It appears that this approach leads to an ~~exploratory data analysis~~ which has a firm probability base. Consequently the distinction between exploratory and confirmatory data analysis can be regarded as a distinction between confirmatory non-parametric statistical

(continued next page)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

Unclassified

next page

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered) cont. data analysis or modeling, and confirmatory parametric statistical data analysis.

Quantile, quantile-density, density-quantile, and score functions are defined in Section 2, and their fundamental inter-relations are discussed. Transformations to observed data which have specified distributions are studied in Section 3, and formulas are given for their derivatives. Auto-regressive representations of density-quantile functions are introduced in Section 4. Sample quantile functions and their linear functionals are defined in Section 5. Goodness of Fit Tests for location and scale parameter models are introduced in Section 6. Estimators of density-quantile functions are discussed in Section 7. Section 8 considers two examples -- Rayleigh data and Buffalo snowfall. Section 9 discusses theoretical examples of density-quantile functions, and their classification according to tail behavior. Location and scale parameter estimation is discussed in Section 10. Section 11 lists some open research problems for extensions.

Nonparametric Statistical Data Modeling*

by

Emanuel Parzen

1. Introduction

"To unlock the analysis of a body of data, to find the good way or ways to approach it, may require a key whose finding is a creative act." writes John Tukey (1977) in the Preface to his book Exploratory Data Analysis. It is the aim of this paper to introduce new types of keys for exploratory data analysis (of continuous data) based on estimating the quantile function and density quantile function. It appears that this approach *leads to an exploratory data analysis which* has a firm probability base. Consequently the distinction between exploratory and confirmatory data analysis can be regarded as a distinction between confirmatory non-parametric statistical data analysis or modeling, and confirmatory parametric statistical data analysis.

Quantile, quantile-density, density-quantile, and score functions are defined in Section 2, and their fundamental inter-relations are discussed. Transformations to observed data which have specified distributions are studied in Section 3, and formulas are given for their derivatives. Autoregressive representations of density-quantile functions are introduced in Section 4. Sample quantile functions and their linear functionals are defined in Section 5. Goodness of Fit Tests for location and scale parameter

* Research supported in part by the Army Research Office (Grant DA AG29-76-0239).

models are introduced in Section 6. Estimators of density-quantile functions are discussed in Section 7. Section 8 considers two examples — Rayleigh data and Buffalo snowfall. Section 9 discusses theoretical examples of density-quantile functions, and their classification according to tail behavior. Location and scale parameter estimation is discussed in Section 10. Section 11 lists some open research problems for extensions.

2. Quantile functions and density-quantile functions

The distribution function (d.f.) of a random variable X is widely denoted $F(x) = P[X \le x]$. The random variable X is said to be continuous (more precisely, absolutely continuous) when F has a probability density function (p.d.f.) $f(x) = F'(x)$ in terms of which

$$F(x) = \int_{-\infty}^{x} f(y)\, dy \quad .$$

Statistical inference has as one of its major aims the estimation (by estimators which are "efficient" or Bayesian, etc.) of $F(x)$ and $f(x)$ from data $X_1,\ldots,X_n$ assumed to be a random sample of X (that is, independent random variables identically distributed as X , denoted i.i.d.).

Parametric statistical inference assumes a representation for $F(x)$ and $f(x)$ as functions of a finite number of parameters, and the estimation problem is posed as one of estimating these parameters. An important parametrization, called the location and scale parameter model, assumes a representation

$$F(x) = F_0\left(\frac{x-\mu}{\sigma}\right) \quad ,$$

$$f(x) = \frac{1}{\sigma} f_0\left(\frac{x-\mu}{\sigma}\right)$$

where F_0 is a specified d.f. and μ and σ are parameters to be estimated (called location and scale parameters respectively).

The procedures for non-parametrically estimating $f(x)$, and for estimating μ and σ , to be introduced in this paper, begin by estimating functions with the following Definitions:

Quantile function (q.f.)	$Q(u) = F^{-1}(u)$, $0 \le u \le 1$
Quantile-density function (q.d.f.)	$q(u) = Q'(u)$, $0 < u < 1$
Density-quantile function (d.q.f.)	$fQ(u) = f\big(Q(u)\big)$, $0 \le u \le 1$
Score function (sc.f.)	$J(u) = -(fQ)'(u)$, $0 < u < 1$.

These functions arise constantly in non-parametric statistics, but they do not seem to be usually given names, or have a universally accepted notation, or be systematically tabulated or discussed (see Hajek and Sidak (1967)).

It is customary mathematical notation to denote a composite function such as $f\big(Q(u)\big)$ by $fQ(u)$; we pronounce it the "eff-cue" function.

For a general distribution function $F(\cdot)$ which is only assumed to be continuous from the right one defines

$$Q(u) = F^{-1}(u) = \inf \{x : F(x) \ge u\} \ .$$

Properties of Q and F can be deduced from each other, using the following fundamental Theorem: for all x in $-\infty < x < \infty$ and all u in $0 < u < 1$

$$F(x) \ge u \text{ if, and only if, } Q(u) \le x \ .$$

(for a proof see Roussas (1973), p. 186 where in addition it is shown that $FQ(u) \ge u$ for any distribution function F).

<u>Theorem</u>. When F is continuous, Q satisfies

$$FQ(u) = u \quad .$$

When F is continuous and strictly increasing there is exactly one x such that $F(x) = u$; then $Q(u)$ equals this value of x and

$$QF(x) = x$$

Differentiating $FQ(u) = u$, we obtain (by the rules for differentiating composite functions) the <u>Reciprocal Theorem</u>:

$$fQ(u)\ q(u) = 1 \quad .$$

In words, fQ and q are reciprocals of each other (which justifies calling them by names which are the reverses of each other).

The q.d. function $q(u)$ thus plays a pivotal role. From a knowledge (or estimator) of q one obtains both $Q(u)$ and $fQ(u)$ by the formulas

$$Q(u) - Q(u_0) = \int_{u_0}^{u} q(t)\, dt \ ,$$

$$fQ(u) = \frac{1}{q(u)} \quad .$$

By the rules for differentiation of composite functions

$$(fQ)'(u) = f'Q(u)\ q(u)$$

so the score function satisfies

$$J(u) = -\frac{f'Q(u)}{f\,Q(u)} = -\frac{f'(F^{-1}(u))}{f\,(F^{-1}(u))}$$

which is the customary definition of the score function in the literature of non-parametric statistics. For purposes of estimation of the score function in small samples, the usual definition requires one to first estimate f', f, and F^{-1}; our definition requires one only to estimate $(fQ)'$ which we will be able to do by a polynomial in $e^{2\pi iu}$.

Many formulas of statistical theory become unified when expressed in terms of quantile functions, density quantile functions, and score functions. The different kinds of <u>tail</u> behavior of distributions clearly correspond to the behavior of $Q(u)$ as u tends to 1 or 0. The formula defining the Pearson family of frequency curves, which is of the form (see Elderton and Johnson (1969))

$$\frac{-f'(x)}{f(x)} = \frac{a_0 + a_1 x}{b_0 + b_1 x + b_2 x^2},$$

can be rewritten, by letting $x = Q(u)$, as a relation between $J(u)$ and $Q(u)$:

$$J(u) = \frac{a_0 + a_1 Q(u)}{b_0 + b_1 Q(u) + b_2 Q^2(u)}$$

Expectations can be expressed in terms of quantile functions. For any function g for which the integrals are finite we have the <u>Theorem</u>:

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)\, f(x)\, dx = \int_0^1 gQ(u)\, du$$

To prove this formula, make the change of variables $x = Q(u)$, $u = F(x)$, $du = f(x)\, dx$. In particular, moments are given by

$$\mu = E(X) = \int_0^1 Q(u)\, du ,$$

$$E(X^2) = \int_0^1 Q^2(u)\, du$$

$$\sigma^2 = \mathrm{Var}\,(X) = \int_0^1 |Q(u) - \mu|^2\, du .$$

We obtain conditions for the integrability of $fQ(u)$ and $\log fQ(u)$ from the Theorem:

$$\int_0^1 g\Big(fQ(u)\Big)\, du = \int_{-\infty}^{\infty} g\Big(f(x)\Big)\, f(x)\, dx$$

whence

$$\int_0^1 fQ(u)\, du = \int_{-\infty}^{\infty} f^2(x)\, dx$$

$$\int_0^1 \log fQ(u)\, du = \int_{-\infty}^{\infty} f(x) \log f(x)\, dx$$

The right hand integrals are familiar in statistical theory, and we believe it is because they are evaluations of the integrals of fQ and $\log fQ$.

The reader interested in examples of fQ functions should see Section 9.

3. Transformations

A basic technique of statistical data analysis, and also of statistical distribution theory, is to transform a continuous random variable X to a continuous random variable $Y = g(X)$ where g is an increasing continuous function. To express the distribution function F_Y of Y in terms of the distribution function F_X of X we have the Theorem:

$$Y = g(X) \text{ implies } F_Y(y) = F_X\left(g^{-1}(y)\right)$$

However the quantile functions are more explicitly related; under the assumption that F_X is a strictly increasing continuous distribution function, we have the Theorem:

$$Y = g(X) \text{ implies } Q_Y(u) = g\left(Q_X(u)\right) \tag{1}$$

which can be deduced from the fact that

$$F_Y(y) \geq u \text{ iff } F_X\left(g^{-1}(y)\right) \geq u \text{ iff } g^{-1}(y) \geq Q_X(u) \text{ iff } y \geq gQ_X(u) .$$

Two important Corollaries are: (i) $Y = \mu + \sigma X$, where $\sigma > 0$, has quantile function $Q_Y(u) = \mu + \sigma Q_X(u)$; (ii) for X positive, $Y = \log X$ has

quantile function	$Q_Y(u) = \log Q_X(u)$
density-quantile function	$f_Y Q_Y(u) = f_X Q_X(u)\ Q_X(u)$
score function	$J_Y(u) = Q_X(u)\ J_X(u) - 1$.

Since a scale parameter can be converted to a location parameter by taking logarithms, it is not surprising that the function $Q_X(u)\ J_X(u) - 1$ arises often in the study of location and scale parameters.

One should keep handy a table of quantile functions of familiar probability laws. If the quantile function of X can be transformed to the quantile function of Y by an increasing continuous transformation g, then to transform X to data identically distributed as Y, form $g(X)$. By perusing a table of quantile functions one immediately obtains the following Theorems (where $Q_0(u)$ appears in parentheses):

(i) log X is extreme value distributed $\left(\log\log \frac{1}{1-u}\right)$ if X is exponential $\left(\log \frac{1}{1-u}\right)$ or Weibull $\left(\left\{\log \frac{1}{1-u}\right\}^{\beta}\right)$;

(ii) if log X is exponential, or log log X is extreme value, then X is Pareto $\left(\{1-u\}^{-\beta}\right)$.

The probability density functions of these distributions is recalled in Section 9.

To simulate a continuous random variable X, one starts with U which is uniformly distributed on 0 to 1 and seeks an increasing function Ψ_1 such that $\Psi_1(U)$ and X are identically distributed; (1) implies that $\Psi_1(u) = Q_X(u)$ as is well known.

Our aim in this paper is to show how to estimate from data increasing functions Ψ and Ψ_1, such that

$$\Psi(X) \sim Y\ , \quad \Psi_1(Y) \sim X$$

where $\sim$ means identically distributed as.

When an observed random variable X is not normal (or exponential) one seeks to find a transformation of data which is normal (or exponential). The cumulative hazard function $H(x)$ in reliability theory, defined by

$$H(x) = -\log\left(1 - F_X(x)\right) ,$$

has the property that $H(X)$ is exponential with mean 1. Thus estimating $H(x)$ can be regarded as actually estimating a transformation to exponentiality.

We are thus led to consider the problem of estimating the transformation Ψ such that $\Psi(X)$ has a prescribed distribution function F_0; further, let Ψ_1 be the transformation such that $\Psi_1(Y) \sim X$ where Y is a random variable with d.f. F_0. Using suitable axioms that Ψ and Ψ_1 be monotone functions, one could prove

$$\Psi(x) = Q_0 F(x) , \quad \Psi_1(y) = Q F_0(y)$$

where F and Q denote the d.f. and q.f. of X. We define these to be the transformations desired since clearly

$$Q_0 F(X) \sim Y , \quad Q F_0(Y) \sim X$$

To find Ψ and Ψ_1 we will find their derivatives

$$\psi(x) = \Psi'(x) , \quad \psi_1(y) = \Psi_1'(y) .$$

The definitions $\Psi = Q_0 F$ and $\Psi_1 = QF_0$ imply the Theorem:

$$\psi(x) = q_0\Big(F(x)\Big)\ f(x)$$

$$\psi_1(y) = q\Big(F_0(y)\Big)\ f_0(y)$$

Now let $x = Q(u)$ and $y = Q_0(u)$; we obtain the Theorem:

$$\psi Q(u) = q_0(u)\ fQ(u)\ ,$$

$$\psi_1 Q_0(u) = q(u)\ f_0 Q_0(u)\ .$$

One immediate conclusion is that ψQ and $\psi_1 Q_0$ are reciprocal functions, so estimating one immediately yields the other.

A second conclusion is that estimating ψQ and estimating fQ are equivalent problems since $f_0 Q_0$ is a known function.

The function which turns out to be natural to estimate is denoted $d(u)$, to indicate that it is a density, with the Definition:

$$d(u) = \frac{1}{\sigma_0} f_0 Q_0(u)\ q(u)\ ,$$

where σ_0 is a normalizing constant with the Definition:

$$\sigma_0 = \int_0^1 f_0 Q_0(u)\ q(u)\ du\ .$$

Conditions for σ_0 to be finite are easily obtained from our general classifaction of fQ functions. σ_0 can be regarded as a scale parameter; its relationship to other measures of scale will be derived from the Theorem (which follows by integration by parts):

$$\sigma_0 = \int_0^1 J_0(u)\ Q(u)\ du$$

assuming $f_0Q_0(u)\ Q(u) = 0$ for $u = 0,1$.

We find it convenient to introduce the following terminology and

Definitions: $d(u)$ is the f_0Q_0 - transformation density of X ,

$$D(u) = \int_0^u d(t)\ dt\ , \quad 0 \leq u \leq 1\ ,$$

is the f_0Q_0 - transformation distribution function of X , and

$$\phi(v) = \int_0^1 e^{2\pi iuv}\ d(u)\ du\ , \quad v = 0, \pm 1, \ldots$$

is the f_0Q_0 - transformation correlation function of X .

A distribution function equal to $\sigma_0 D(u)$ has been extensively studied in reliability theory (see Barlow and Doksum (1972)) under the notation

$$H_F^{-1}(u) = \int_0^{F^{-1}(u)} f_0[F_0^{-1}F(x)]\ dx$$

which we write in our notation, letting $t = F(x)$,

$$H_F^{-1}(u) = \int_0^u f_0Q_0(t)\ q(t)\ dt$$

What is novel in our approach is that we consider the density function and Fourier transform of this distribution function.

Recently, Barlow and Campo (1975) and Barlow and Proschan (1977) have studied the statistic

$$H_F^{-1}(u) = \int_0^{F^{-1}(u)} \{1 - F(x)\}\, dx$$

which they call the <u>total time on test transform</u> of the distribution F , and use it to test for exponentiality. It is the same as our $\sigma_0 D(u)$ with $f_0 Q_0 = 1-u$, the density-quantile function of the exponential distribution.

4. Density-Quantile Autoregressive Representations as Generalizations of Goodness of Fit Hypotheses

The concepts have now been defined to state our new approach to statistical data analysis. Given a random sample $X_1, \ldots, X_n$ of a random variable X one would like to test the hypothesis H_0 that the data is normal (or exponential or any other specified type) and/or one would like to find a transformation of the data after which it is normal (or exponential or any other specified type). By a specified type we mean that the true d.f. F is of the location-scale parameter form

$$F(x) = F_0\left(\frac{x - \mu}{\sigma}\right)$$

where μ and σ are parameters to be efficiently estimated, and F_0 is specified. When testing normality, $F_0(x) = \Phi(x)$, the standard normal distribution function.

Theorem: H_0 is equivalent to any one of the following hypotheses:

$$Q(u) = \mu + \sigma Q_0(u)\,, \quad q(u) = \sigma q_0(u)\,, \quad fQ(u) = \frac{1}{\sigma} f_0 Q_0(u)\,,$$
$$d(u) = 1\,, \quad D(u) = u\,, \quad \phi(v) = 0 \text{ for } v \neq 0\,.$$

When the density $d(u)$ is constant, it is called "white noise" in honor of an analogous situation in time series analysis. An approach to testing this hypothesis which also provides an estimator of $d(u)$ when we do not believe it to be a constant is to represent it in a form called an autoregressive representation (since it is analogous to the spectral density of an autoregressive scheme in time series analysis).

Definition: A density $d(u)$ is said to be autoregressive of order m , or to have an autoregressive representation of order m , if it is of the form

$$d(u) = K_m |1 + \alpha_m(1)\, e^{2\pi iu} + \ldots + \alpha_m(m)\, e^{2\pi ium}|^{-2} \qquad (1)$$

where m is an integer called the <u>order</u> (whose determination is the most difficult estimation problem), K_m is a positive constant (corresponding to the finite memory m one-step ahead mean square prediction error), and $\alpha_m(1),\ldots,\alpha_m(m)$ are <u>complex-valued</u> coefficients satisfying the condition that

$$g_m(z) = 1 + \alpha_m(1)\, z + \ldots + \alpha_m(m)\, z^m$$

has all its roots outside the unit circle. (For future reference note that z^* denotes the complex conjugate of z).

When $d(u) = \dfrac{f_0Q_0(u)}{\sigma_0 fQ(u)}$ is autoregressive of order m , one obtains a representation for fQ which generalizes the formula which holds in the location and scale parameter model:

$$fQ(u) = c_m \left|1 + \alpha_m(1)\, e^{2\pi iu} + \ldots + \alpha_m(m)\, e^{2\pi ium}\right|^2 f_0Q_0(u) \qquad (2)$$

where

$$\frac{1}{c_m} = \int_0^1 |1 + \alpha_m(1)\, e^{2\pi iu} + \ldots + \alpha_m(m)\, e^{2\pi ium}|^2 f_0Q_0(u)\, q(u)\, du$$

In fact we use low order schemes to represent d(u) . We thus consider successively representations for fQ(u) of the form

<u>m = 0</u> $\quad fQ(u) = c_0\, f_0Q_0(u)$

<u>m = 1</u> $\quad fQ(u) = c_1\, |1 + \alpha_1(1)\, e^{2\pi iu}|^2\, f_0Q_0(u)$

<u>m = 2</u> $\quad fQ(u) = c_2\, |1 + \alpha_2(1)\, e^{2\pi iu} + \alpha_2(2)\, e^{2\pi iu2}|^2\, f_0Q_0(u)$

and so on. It is clear that we have a sequence of representations for fQ which start with the hypothesis H_0 and ascend to the general representation

$$fQ(u) = f_0Q_0(u)\; c_\infty \left|1 + \alpha_\infty(1)\, e^{2\pi iu} + \dots + \alpha_\infty(m)\, e^{2\pi ium} + \dots\right|^2 \qquad (3)$$

The infinite-order autoregressive representation (3) holds when conditions such as the following are true (see Geronimus (1960)): first,

$$\frac{fQ(u)}{f_0Q_0(u)}\,, \quad \frac{f_0Q_0(u)}{fQ(u)}\,, \quad \log fQ(u)\,, \quad \log f_0Q_0(u)$$

are all integrable over $0 \le u \le 1$; second, fQ and f_0Q_0 satisfy a smoothness condition such as differentiability. The speed of convergence of the approximations of order m to the infinite order case depends on the number of derivatives that exist, and is exponentially fast for infinitely differentiable functions.

<u>Theorem</u>: The coefficients of an autoregressive representation of order m for the f_0Q_0 - transformation density $d(u)$ can be computed from a knowledge of the f_0Q_0 - transformation correlations $\phi(0),\phi(1),\phi(-1),\dots,\phi(m),\phi(-m)$ up to lag m using the difference equation satisfied by $\phi(v)$:

$$\phi(-v) + \alpha_m(1)\, \phi(1 - v) + \dots + \alpha_m(m)\, \phi(m - v) = 0\,, \quad v > 0$$

$$\phi(0) + \alpha_m(1)\, \phi(1) + \dots + \alpha_m(m)\, \phi(m) = K_m\;.$$

<u>Proof</u>: Since $d(u) = K_m\{g_m(e^{2\pi iu})(g_m(e^{2\pi iu}))^*\}^{-1}$ we can write

$$\phi(-v) + \alpha_m(1)\ \phi(1 - v) + \ldots + \alpha_m(m)\ \phi(m - v)$$

$$= \int_0^1 e^{-2\pi iuv}\ g_m(e^{2\pi iu})\ d(u)\ du$$

$$= \int_0^1 e^{-2\pi iuv}\ K_m\left\{\left(g_m(e^{2\pi iu})\right)^*\right\}^{-1} du$$

Now $\left(g_m(e^{2\pi iu})\right)^*$ is a polynomial in $e^{-2\pi iu}$ whose reciprocal has a convergent power series in positive powers of $e^{-2\pi iu}$ (with constant term equal to 1) by virtue of the assumption on the location of the zeroes of $g(z)$. Since $\int_0^1 e^{-2\pi iu(v+k)}\ du = 0$ for positive v and k, the above expression equals 0 for $v > 0$, and equals K_m for $v = 0$.

5. Sample Quantile Function

Given a sample $X_1,\ldots,X_n$ of a continuous random variable X, we denote the empirical distribution function (EDF) by $\tilde{F}(x)$, read F wiggle; it is defined by

$$\tilde{F}(x) = \text{fraction of } X_1,\ldots,X_n \le x \quad .$$

We shall give several definitions of the empirical quantile function (EQF) denoted $\tilde{Q}(u)$. The first definition is

$$\tilde{Q}(u) = \tilde{F}^{-1}(u) = \inf \{x : \tilde{F}(x) \ge u\} \quad .$$

It is a piecewise constant function whose values are the order statistics $X_{(1)} < X_{(2)} < \ldots < X_{(n)}$; more precisely,

$$\tilde{Q}(u) = X_{(j)} \quad \text{for} \quad \frac{j-1}{n} < u \le \frac{j}{n}, \quad j = 1,\ldots,n \quad .$$

For $u = 0$, we define $\tilde{Q}(0) = X_{(0)}$ where $X_{(0)}$ is taken to be either the sample minimum $X_{(1)}$ or a natural minimum when one is available (when X is non-negative, one might take $X_{(0)} = 0$).

If one desires to form a smooth function from a wiggly function, it seems reasonable to start with the smoothest reasonable definition (which is differentiable if possible). Consequently a preferable definition of $\tilde{Q}(u)$ might be the piecewise linear function

$$\tilde{Q}(u) = n\left(\frac{j}{n} - u\right) X_{(j-1)} + n\left(u - \frac{j-1}{n}\right) X_{(j)}$$

$$\text{for } \frac{j-1}{n} \le u \le \frac{j}{n} \text{ and } j = 1,\ldots,n \ .$$

Then $\tilde{q}(u) = \tilde{Q}'(u)$ is given by

$$\tilde{q}(u) = n(X_{(j)} - X_{(j-1)})$$

$$\text{for } \frac{j-1}{n} < u < \frac{j}{n} \text{ and } j = 1,\ldots,n \ .$$

We call $n(X_{(j)} - X_{(j-1)})$, $j = 1,\ldots,n$ the <u>spacings</u> of the sample (see Pyke (1965), (1972)). The most important fact about $\tilde{q}(u)$ is that it is asymptotically exponentially distributed with mean $q(u)$. The sample spectral density of a stationary time series has an analogous property. Consequently there is an <u>isomorphism</u> between spacings and sample spectral densities; to any result about one there is an analogous result about the other. The methods of proofs and exact hypotheses may need to be different for the two cases, but the statement of the conclusion is usually found to be the same.

Estimators which may have better behavior in small samples from symmetric densities can be obtained by adopting a shifted piecewise linear function as the definition of $\tilde{Q}(u)$:

$$\tilde{Q}(u) = n\left(\frac{2j+1}{n} - u\right) X_{(j)} + n\left(u - \frac{2j-1}{2n}\right) X_{(j+1)}$$

$$\text{for } \frac{2j-1}{n} \le u \le \frac{2j+1}{2n} \quad \text{and} \quad j = 1,\ldots,n-1 \;;$$

$$= \text{undefined for} \quad u < \frac{1}{2n} \quad \text{or} \quad u > 1 - \frac{1}{2n} \;.$$

Its derivative is

$$\tilde{q}(u) = n(X_{(j+1)} - X_{(j)}) \,, \quad \frac{2j-1}{2n} < u < \frac{2j+1}{2n} \;.$$

Finally a Bayesian definition of $\tilde{Q}(u)$ can be adopted, using the Fractional Order Statistics Process defined by Stigler (1977).

For plotting of sample quantile functions, we have found it useful to normalize them:

$$\bar{\tilde{Q}}(u) = \frac{\tilde{Q}(u) - \tilde{Q}(0)}{\tilde{Q}(1) - \tilde{Q}(0)}$$

This is a monotone function on $0 \le u \le 1$ whose values lie between 0 and 1 . In my view, normalized graphs enable one to apply the experience obtained in analyzing data of one kind to the analysis of data of another kind.

The asymptotic distribution of the quantile process $\tilde{Q}(u)$, $0 \le u \le 1$, is usually studied in the literature for the first definition; the work most useful to us is that of Czorgo and Revesz (1975), (1978) described in Sections 9 and 10 of this paper (see also Shorack (1972)). An open research problem is to show that this asymptotic distribution theory applies also to the other definitions of $\tilde{Q}$ we have given.

The basic estimators we form in practice are linear functionals $T = \int_0^1 W(u)\, d\tilde{Q}(u)$. For the first definition of $\tilde{Q}$,

$$T = \sum_{j=0}^{n-1} W\left(\frac{j}{n}\right)\{X_{(j+1)} - X_{(j)}\}$$

For the second definition of $\tilde{Q}$,

$$T = \sum_{j=1}^{n} n(X_{(j)} - X_{(j-1)}) \int_{(j-1)/n}^{j/n} W(u)\, du$$

We might evaluate the integral by a simple Simpson's rule approximation:

$$\int_{(j-1)/n}^{j/n} W(u)\, du = \frac{1}{6}\left\{W\left(\frac{j-1}{n}\right) + 4W\left(\frac{2j-1}{2n}\right) + W\left(\frac{j}{n}\right)\right\}$$

For the third definition of $\tilde{Q}$,

$$T = \sum_{j=1}^{n-1} n(X_{(j+1)} - X_{(j)}) \int_{(2j-1)/2n}^{(2j+1)/2n} W(u)\, du$$

When $W(u) = e^{2\pi iuv} f_0Q_0(u)$, we might approximate the last integral by $f_0Q_0\left(\frac{j}{n}\right) e^{2\pi iv(j/n)} \frac{\sin(\pi v/n)}{\pi v}$.

The distribution theory of linear functions of order statistics has an extensive literature (see Chernoff, Gastwirth, and Johns (1967), Moore (1968) , Stigler (1974)).

6. Goodness of Fit Tests

Given a sample $X_1,\ldots,X_n$ of a continuous random variable X, one forms the EQF $\tilde{Q}(u)$ and empirical quantile-density function $\tilde{q}(u)$. Then for each probability law type whose goodness of fit one might want to test, there is a corresponding standard f_0Q_0 function. For each specified f_0Q_0 function one would compute:

I. Sample Transformation-Density Function or Weighted Spacings

$$\tilde{d}(u) = \frac{1}{\tilde{\sigma}_0} f_0Q_0(u)\,\tilde{q}(u)$$

$$\tilde{\sigma}_0 = \int_0^1 f_0Q_0(u)\,\tilde{q}(u)\,du \quad .$$

II. Sample Transformation-Distribution Function or Cumulative Weighted Spacings

$$\tilde{D}(u) = \int_0^u \tilde{d}(t)\,dt\ , \quad 0 \le u \le 1\ .$$

III. Sample Transformation Correlations

$$\tilde{\phi}(v) = \int_0^1 e^{2\pi i u v}\,\tilde{d}(u)\,du\ , \quad v = 0, \pm 1, \pm 2, \ldots \quad .$$

To test the Goodness of Fit Hypothesis H_0 one has available test statistics as follows:

I. $$\max_{0 \le u \le 1} \tilde{d}(u)\ , \quad \int_0^1 \log \tilde{d}(u)\,du \quad .$$

II. $\tilde{D}(0.5)$, $\tilde{D}(.75) - \tilde{D}(.25)$

$$\max_{0 \le u \le 1} |\tilde{D}(u) - u| \ , \quad \int_0^1 |\tilde{D}(u) - u|^2 \, du$$

$$\int_0^1 J_1(u) \, d\tilde{D}(u) \quad \text{for a specified } J_1(u)$$

III. sequence $|\tilde{\phi}(1)|^2, |\tilde{\phi}(2)|^2, \ldots$

$$\sum_{v \neq 0} k(v) \, |\tilde{\phi}(v)|^2 \quad \text{for a specified } k(v) \quad .$$

The distribution theory of many of these statistics have already been studied in the literature. For a general $D(u)$, the almost sure convergence to 0 of $\max_{0 \le u \le 1} |\tilde{D}(u) - D(u)|$ was proved by Barlow and van Zwet (1970). The asymptotic distribution of $\tilde{\sigma}_0$ was found by Weiss (1964). The asymptotic distribution of $\int_0^1 \log \tilde{d}(u) \, du$ is the same as that of the sample innovation variance, as given by Davis and Jones (1968) and Hannan and Nicholls (1977).

Under H_0 , the asymptotic distribution of $\max_{0 \le u \le 1} \tilde{d}(u)$ is the same as the distribution in time series analysis (first found by Fisher (1929)) of the maximum normalized periodogram ordinate of white noise.

An important open research problem is the following Conjecture: under H_0 , the stochastic process $\sqrt{n} \, \{\tilde{D}(u) - u\}$, $0 \le u \le 1$ is asymptotically distributed as a Brownian Bridge process $B(u)$, $0 \le u \le 1$;

this has been proved for $f_0Q_0(u) = 1 - u$, corresponding to the exponential distribution (Barlow (1976) personal communication). It would then follow that all statistics based on $\tilde{D}(u) - u$ have the same asymptotic distribution theory as the corresponding statistics based on $\tilde{F}(x) - x$, $0 \le x \le 1$, where $\tilde{F}(x)$ is the EDF of a random sample from a uniform distribution on $[0,1]$ whose theory is summarized by Durbin (1973).

The foregoing framework includes as special cases many goodness of fit test statistics that are being proposed (for example, Andrews' test for normality (Gnandesikan (1977), p. 165) and tests for Weibull and extreme value distributions introduced by Mann and Fertig (1975)).

In the next section we propose additional Goodness of Fit Tests based on determining the order of an autoregressive smoother to $\tilde{d}(u)$.

7. Density-Quantile Autoregressive Estimation

Given a sample $X_1,\ldots,X_n$ of a continuous random variable, we have discussed how to test a Goodness of Fit Hypothesis by forming the sample functions $\tilde{d}(u)$, $\tilde{D}(u)$, and $\tilde{\phi}(v)$. In this section we discuss how to form autoregressive densities of order m, $\hat{d}_m(u)$, $m = 0,1,\ldots$ which are candidates for estimators of the true density $d(u)$. The sequence has the property that $\hat{d}_0(u)$ is constant (identically equal to 1) and $\hat{d}_m(u)$ tends to $\tilde{d}(u)$ as m increases.

For time series spectral estimation by autoregressive estimators Parzen (1974), (1977) has introduced a criterion called CAT (criterion autoregressive transfer function) for determining the optimal order $\hat{m}$ such that $\hat{d}_{\hat{m}}(u)$ is an optimal estimator of $d(u)$. We calculate an analogous criterion for smooth densities $\hat{d}_m(u)$, defined by

$$\text{CAT}(m) = \frac{1}{n}\sum_{j=1}^{m} \hat{K}_j^{-1} - \hat{K}_m^{-1} .$$

The distribution theory of CAT(m) is known approximately only under H_0. Consequently, at the present time we regard CAT as interpretable only when it chooses $\hat{m} = 0$; then we regard it as additional confirmation that H_0 holds (when this hypothesis is accepted by tests based on $\tilde{d}$, $\tilde{D}$, and/or $|\tilde{\phi}|^2$).

A graphical approach to choosing the appropriate smooth estimator $\hat{d}_m(u)$ which is the most likely estimator of $d(u)$ is to use as a criterion how $\hat{D}_m(u) = \int_0^u \hat{d}_m(t)\,dt$ fits $\tilde{D}(u)$. If it fits too well one has over-smoothed, and the density $\hat{d}_m(u)$ will have spurious modes. One wants $\hat{D}_m(u)$ to follow $\tilde{D}(u)$ but not slavishly.

Next we define the autoregressive estimators and state a theorem concerning their consistency.

The autoregressive smoother of order m, denoted $\hat{d}_m(u)$, is defined to be

$$\hat{d}_m(u) = \hat{K}_m \left|1 + \hat{\alpha}_m(1)\, e^{2\pi iu} + \ldots + \hat{\alpha}_m(m)\, e^{2\pi ium}\right|^{-2} ,$$

where $\hat{\alpha}_m(1),\ldots,\hat{\alpha}_m(m)$ are the values of $\alpha_m(1),\ldots,\alpha_m(m)$ minimizing

$$\int_0^1 |g_m(e^{2\pi iu})|^2\, \tilde{d}(u)\, du ,$$

where $g_m(z) = 1 + \alpha_m(1)\, z + \ldots. + \alpha_m(m)\, z^m$;

$$\hat{K}_m = \int_0^1 |\hat{g}_m(e^{2\pi iu})|^2\, \tilde{d}(u)\, du$$

where $\hat{g}_m(z) = 1 + \hat{\alpha}_m(1)\, z + \ldots + \hat{\alpha}_m(m)\, z^m$.

By the projection theorem in Hilbert space, $\hat{g}_m(z)$ satisfies the orthogonality conditions

$$\int_0^1 \hat{g}_m(e^{2\pi iu})\, e^{-2\pi iuv}\, \tilde{d}(u)\, du = 0$$

for $v = 1,\ldots,m$ which is equivalent to the normal equations

$$\tilde{\phi}(-v) + \hat{\alpha}_m(1)\, \tilde{\phi}(1 - v) + \ldots + \hat{\alpha}_m(m)\, \tilde{\phi}(m - v) = 0 \qquad (1)$$

for $v = 1,\ldots,m$. Next, the orthogonality conditions imply

$$\hat{K}_m = \int_0^1 \hat{g}_m(e^{2\pi iu}) \, \tilde{d}(u) \, du$$

$$= 1 + \hat{\alpha}_m(1) \, \tilde{\phi}(1) + \ldots. + \hat{\alpha}_m(m) \, \tilde{\phi}(m) \; .$$

A rigorous theorem concerning the consistency in probability of $\hat{d}_m(u)$ as an estimator of $d(u)$ can be proved by adapting the work of Carmichael (1976) in his Ph.D. thesis on the autoregressive method for probability density estimation.

<u>Theorem</u> (Carmichael (1976)). If

(1) $d(u)$, $d^{-1}(u)$, $\log d(u)$ are integrable

(2) $d(u)$ is bounded above and below in the sense

$$0 < d_L \le d(u) \le d_U < \infty \qquad \text{a.e. in } [0,1]$$

(3) $d(u) = c(u)$ a.e. $[0,1]$ and c satisfies, for some $\alpha > 0.5$,

$$\sup_{|h|\le\delta} \int_0^1 |c(u + h) - c(u)|^2 \, du = O(\delta^{2\alpha})$$

(4) m is chosen as a function of the sample size n satisfying

$$\lim_{n\to\infty} \frac{m^3}{n} = 0$$

then as $n \to \infty$

$$\sup_{0\le u\le 1} |\hat{d}_m(u) - d_a(u)| \to 0 \text{ in probability}$$

where $d_a(u)$ is a density function with an infinite autoregressive representation

$$d_a(u) = K_\infty \left| g_\infty(e^{2\pi iu}) \right|^{-2}$$

and satisfying $d_a(u) = d(u)$ a.e. in $[0,1]$.

The proof of this beautiful theorem is being submitted for publication.

An estimator $\hat{d}_m(u)$ of $d(u)$ yields an estimator $\hat{fQ}_m(u)$ of fQ which is given explicitly by

$$\hat{fQ}_m(u) = \frac{\left|1 + \hat{\alpha}_m(1)\, e^{2\pi iu} + \ldots + \hat{\alpha}_m(m)\, e^{2\pi ium}\right|^2 f_0Q_0(u)}{\int_0^1 \left|1 + \hat{\alpha}_m(1)\, e^{2\pi iu} + \ldots + \hat{\alpha}_m(m)\, e^{2\pi ium}\right|^2 f_0Q_0(u)\, \tilde{q}(u)\, du}$$

where $\hat{\alpha}_m(1),\ldots,\hat{\alpha}_m(m)$ are the solutions of the normal equations (1).

To compare our autoregressive estimator $\hat{fQ}_m(u)$ with other possible estimators one must realize that we are actually estimating the triple of functions fQ , q , and Q , and the basic aim is to form a smooth function $\hat{Q}$ which is an estimator of Q . One can distinguish three general approaches to forming estimators $\hat{Q}$ which we call

I. Parametric

II. Non-parametric

III. Non-parametric pre-flattened.

The parametric approach assumes a location and scale parameter representation $Q(u) = \mu + \sigma Q_0(u)$, forms efficiently estimators $\hat{\mu}$ and $\hat{\sigma}$, and then takes

$$\hat{Q}(u) = \hat{\mu} + \hat{\sigma} Q_0(u) \text{ as the estimator of } Q .$$

The non-parametric approach estimates Q at a point by averaging over the values of $\tilde{Q}(p)$ for p in a neighborhood of u. An estimator of this form is usually written as a kernel estimator

$$\hat{Q}(u) = \int_0^1 \tilde{Q}(p) \frac{1}{h} K\left(\frac{u-p}{h}\right) dp$$

for a suitable kernel K and bandwidth h. If one adopts the piecewise-linear definition of $\tilde{Q}$, one can differentiate this formula for $\hat{Q}$ to form a smooth estimator $\hat{q}$ of the quantile-density q:

$$\hat{q}(u) = \int_0^1 \tilde{q}(p) \frac{1}{h} K\left(\frac{u-p}{h}\right) dp$$

Estimators of this form are in fact extensively studied in the literature of non-parametric density estimation (see Bofinger (1975), Moore and Yackel (1977)). under the name of "nearest neighbor density estimates." Another approach to fitting smooth curves $\hat{q}$ to the wiggly function $\tilde{q}$ is to use splines (see Wahba and Wold (1975)).

The foregoing estimators of q will have good properties only at a fixed value of u; the consistency of estimation becomes worse as u tends to 0 or 1 because $q(u)$ is in general a non-integrable function. This problem can be overcome by multiplying $q(u)$ by a factor $f_0Q_0(u)$ which makes the product $f_0Q_0(u)\, q(u)$ an integrable function, which is not oscillating as much. When one smooths not $\tilde{q}(u)$ but $f_0Q_0(u)\, \tilde{q}(u)$, we call the approach non-parametric preflattened smoothing. We smooth $\tilde{d}(u) = f_0Q_0(u)\ \tilde{q}(u) \div \tilde{\sigma}_0$. One approach would be to form estimators of the form

$$\hat{d}(u) = \int_0^1 \tilde{d}(p) \frac{1}{h} K\left(\frac{u-p}{h}\right) dp$$

It is difficult to use this approach in practice because of difficulties in optimally choosing h . We believe the autoregressive approach to density estimation goes a long way towards overcoming these difficulties.

For the mathematical statistician, many problems are open for research concerning the asymptotic distributions of the foregoing estimators.

8. Computing Routines and Examples

A computer program which implements the data analysis approach described here has been developed by Prof. J. P. Carmichael and Mr. David Tritchler. Given a sample $X_1,\ldots,X_n$ it: (1) lists their order statistics, means, variances, etc.; (2) plots the normalized quantile function; (3) plots spacings. The f_0Q_0 functions of various familiar probability laws are available to be applied. For a specified f_0Q_0 function, the computer programs (4) plots $\tilde{d}(u)$ the raw transformation-density function; (5) plots $\tilde{D}(u)$, the raw transformation-distribution function; (6) plots $|\tilde{\phi}(v)|^2$, the square-modulus raw transformation-correlations. Next for $m = 1,2,\ldots$, the autoregressive approximator $\hat{d}_m(u)$ is computed, and its distribution function $\hat{D}_m(u)$ is plotted superimposed on a graph of $\tilde{D}(u)$ to enable one to see how well $\hat{D}_m(u)$ fits $\tilde{D}(u)$. Finally, CAT, a criterion to help determine the optimal order m of autoregressive approximation, is tabulated, and the order at which CAT achieves its minimum is determined. In addition, for each m the density-quantile estimator $\hat{fQ}_m(u)$ corresponding to $\hat{d}_m(u)$ is plotted. In the absence of a rigorous procedure for determining the optimal order $\hat{m}$, we choose those values of m for which $\hat{D}_m(u)$ "fits" $\tilde{D}(u)$.

Rayleigh example. Tukey (1977), p. 49, gives an example of data (Rayleigh's weights of a standard volume of "nitrogen" consisting of 15 measurements) which can be used to look hard at the advantages and disadvantages of graphical data analysis techniques. Rayleigh's observations in 1893-1894 established a discrepancy between the densities of nitrogen produced by removing the oxygen from air and nitrogen produced by decomposition of a chemical compound which led him to investigate the composition of air chemically freed of oxygen which led to the discovery of argon, for which Rayleigh (1842-1919) was awarded the 1904 Nobel Prize in Physics.

We may define the goal of statistical data analysis techniques as follows: on the one hand, to enable the scientist to win a Nobel Prize; on the other hand, to protect the statistician from being sued by a scientist who claims that using the statistician's techniques prevented him (her) from winning a Nobel Prize.

Tukey discusses how to present the data so as to make it quite clear that it separates into two quite isolated subgroups, which one interprets as indicating that the single batch of weights might be two batches of weights (as in fact they are, one for "nitrogen" from air, the other for "nitrogen" from other sources).

The presence of two batches will be indicated by the shapes of the empirical quantile function or spacings. However, I believe it is most clearly indicated by the presense of two modes in the estimated density-quantile function. We usually estimate fQ taking as the base function f_0Q_0 the standard normal density, so that the procedure also provides a test of normality. The Rayleigh data is clearly non-normal. We take order $m = 2$ as an optimal autoregressive approximation (on the criterion of the fit of $\hat{D}_2(u)$ to $\tilde{D}(u)$) and obtain the estimated density-quantile function whose plot appears in Figure I; it is bimodal.

The reader may find it interesting to compare the density-quantile function plot in Figure I with Tukey's two batches box and whiskers plot in Tukey (1977), p. 51. Our left hand mode (representing "other than air" nitrogen measurements) is lower than the right hand mode (representing "from air" nitrogen measurements), indicating that the left mode population is more variable than the right mode population.

Buffalo snowfall example. The 63 yearly values of snow precipitation in Buffalo (recorded to the nearest tenth of an inch) from 1910-1972 have been extensively analyzed by Carmichael (1976) and Thaler (1972) to illustrate and compare various probability density estimation techniques. Different analyses have indicated either a uni-modal or tri-modal density, with the tri-modal shape usually regarded as the more likely answer. In our density-quantile estimation procedure, with base f_0Q_0 taken to be the standard normal, the order 0 and order 1 autoregressive estimator $\hat{fQ}_1(u)$ are unimodal, and the order 2 autoregressive estimator $\hat{fQ}_2(u)$ is tri-modal (see Figure II). However all our $\tilde{D}$ and $|\tilde{\phi}|^2$ based diagnostic tests of the hypothesis H_0 that Buffalo snowfall is normal confirm that it is. Thus the trimodal density estimator often obtained in previous analyses seems not to be correct. It is interesting that Tukey (1977), p. 117 also suggests Buffalo snowfall as an example for analyses (and gives the data for 1918-1937).

Figure I. Rayleigh data. Crosses represent cumulative weighted spacings function $\tilde{D}$. Solid line represents autoregressive estimator $\hat{D}_2$ of order 2 .

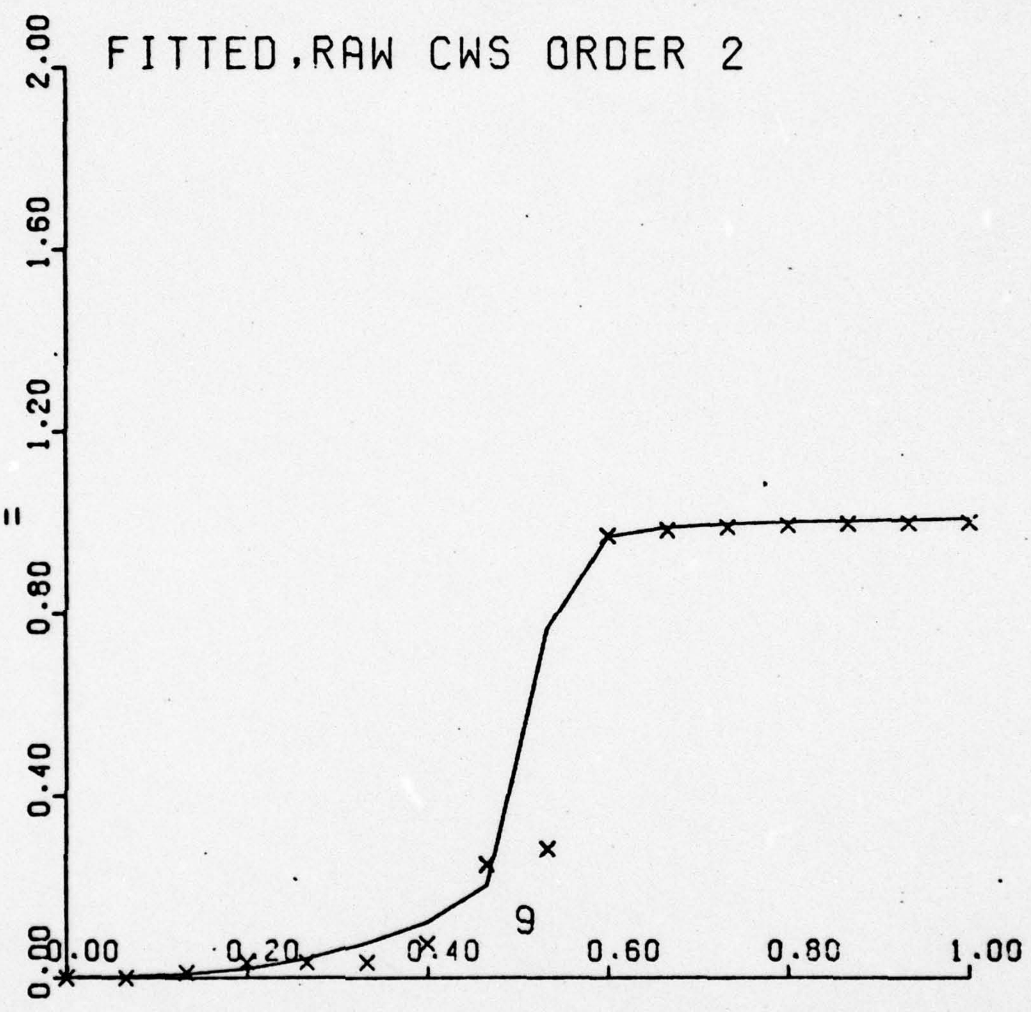
FITTED,RAW CWS ORDER 2
2.00
1.60
1.20
0.80
0.40
0.00
0.00
0.20
0.40
0.60
0.80
1.00
9

Figure I.

Figure II. Rayleigh data. Autoregressive estimator $\hat{fQ}$ of density quantile function. Order 2 chosen on basis of fit of $\hat{D}$ to $\tilde{D}$.

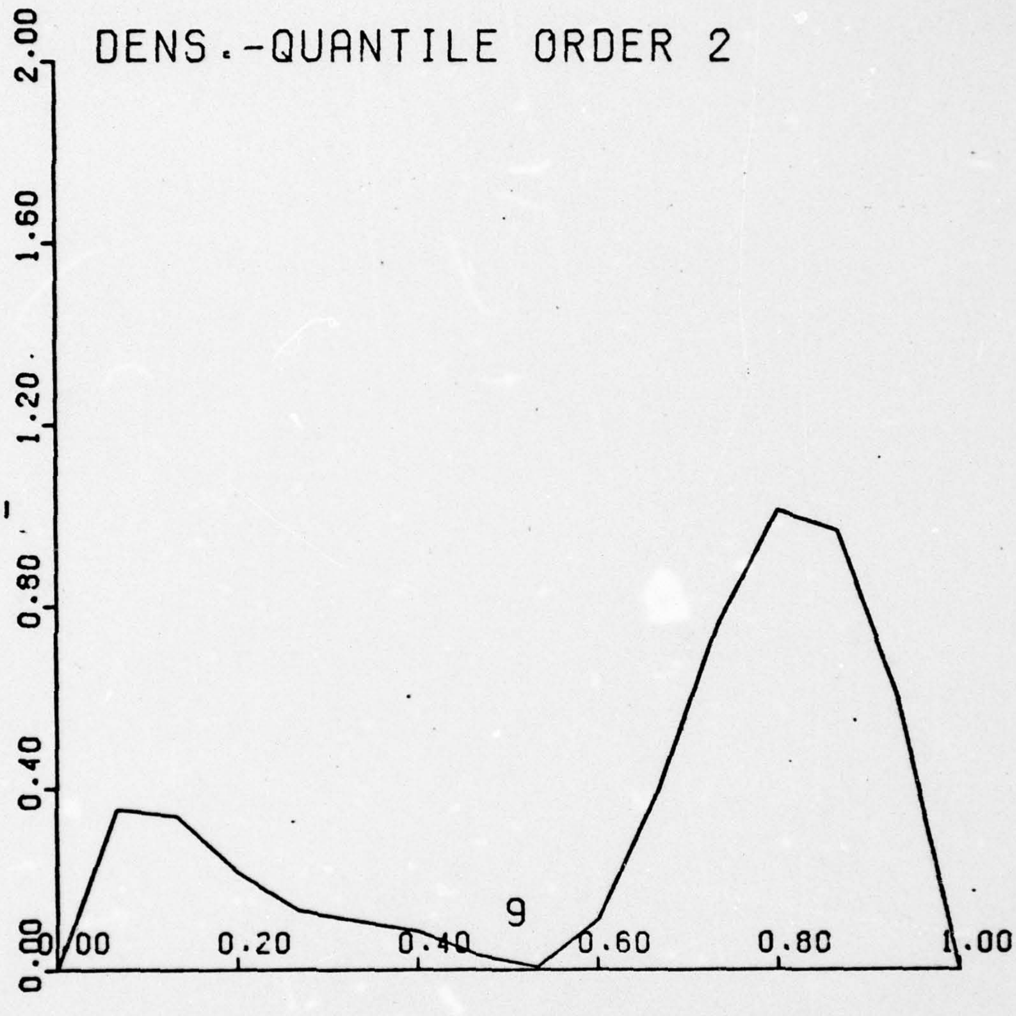
DENS.-QUANTILE ORDER 2
0.00
0.40
0.80
1.20
1.60
2.00
0.00
0.20
0.40
0.60
0.80
1.00
9

Figure II.

Figure III. Buffalo Snowfall Data. The sample quantile function $\tilde{Q}$ is in upper left graph, spacings or sample quantile-density function $\tilde{q}$ is in lower left graph, normal weighted spacings $\tilde{d} = \phi\Phi^{-1}\tilde{q}$ is in lower right graph, and cumulative weighted spacings $\tilde{D}$ is in upper right graph.

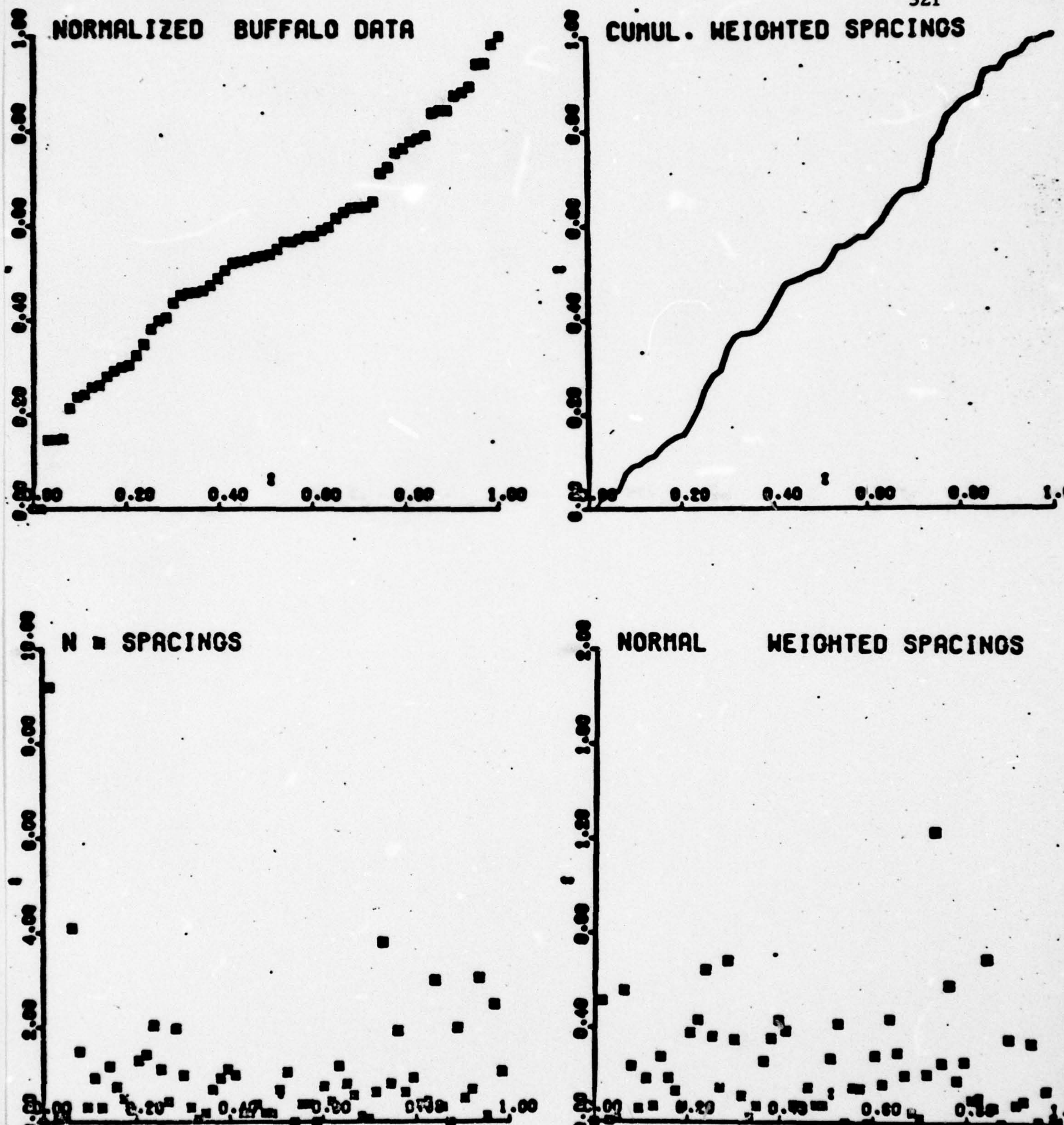
NORMALIZED BUFFALO DATA
CUMUL. WEIGHTED SPACINGS
N = SPACINGS
NORMAL WEIGHTED SPACINGS
0.00
0.20
0.40
0.60
0.80
1.00

Figure III.

Figure IV. Buffalo Snowfall Data. Upper left graph depicts $\tilde{D}$ by crosses and $\hat{D}_1$ by solid line; upper right graph depicts $\tilde{D}$ by crosses and $\hat{D}_2$ by solid line. Autoregressive estimators $\hat{fQ}_1$ and $\hat{fQ}_2$ of orders 1 and 2 appear in lower left and lower right graphs respectively.

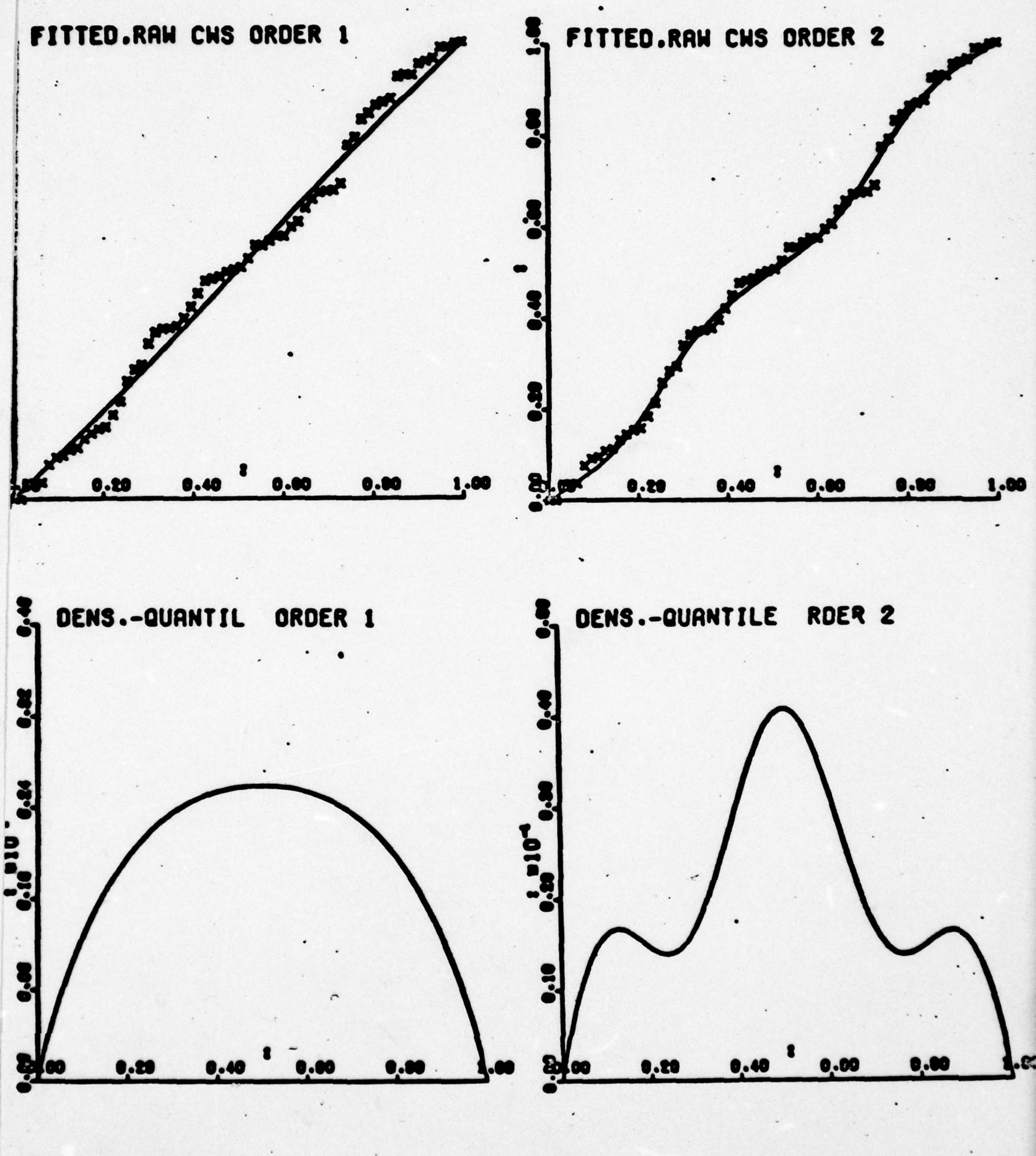
FITTED.RAW CWS ORDER 1
0.20
0.40
0.60
0.80
1.00
FITTED.RAW CWS ORDER 2
DENS.-QUANTIL ORDER 1
DENS.-QUANTILE RDER 2

Figure IV.

9. Density-Quantile Classification of Probability Laws

An examination of the density-quantile functions $fQ(u)$ of familiar probability laws indicate that they can be classified according to their limiting behavior as u tends to 0 or 1. The behavior as $u \to 1$ can be described as either

$$fQ(u) \sim (1-u)^{\alpha} \qquad , \quad \alpha > 0$$

or

$$fQ(u) \sim (1-u)\left\{\log \frac{1}{1-u}\right\}^{1-\beta} \qquad , \quad 0 \le \beta \le 1$$

where $g_1(u) \sim g_2(u)$ means $g_1(u) \div g_2(u)$ tends to a positive finite constant (as $u \to 1$).

We call α the <u>tail-exponent</u> parameter and β the <u>shape</u> parameter of a distribution. A rigorous definition of the tail exponent is given at the end of the section.

The parameter ranges $\alpha < 1$, $\alpha = 1$, and $\alpha > 1$ correspond to the statistician's perception that probability laws have three types of tail behavior:

I. SHORT TAILS OR LIMITED TYPE

II. MEDIUM TAILS OR EXPONENTIAL TYPE

III. LONG TAILS OR CAUCHY TYPE

The names limited type, exponential type, or Cauchy type are used in the theory of extreme value distributions to describe the types of distributions leading to the three types of extreme value distributions (see Gumbel (1962)).

The uniform distribution has $\alpha = 0$:

$$f(x) = 1, \quad 0 \le x \le 1 \quad ; \quad fQ(u) = 1, \quad 0 \le u \le 1$$

An example of a short-tailed distribution is

$$f(x) = c(1 - x)^{c-1}, \quad 0 \le x \le 1 \quad ; \quad fQ(u) = \frac{1}{\beta}(1 - u)^{1-\beta}$$

where $c > 0$ and $\beta = 1/c$.

Examples of exponential distributions are

exponential	$e^{-x}, \; x > 0$;	$fQ(u) = 1 - u$
logistic	$\dfrac{e^x}{(1 + e^x)^2}, \; -\infty < x < \infty$;	$fQ(u) = u(1 - u)$
Weibull $c = \frac{1}{\beta} > 0$	$c\,x^{c-1} e^{-x^c}, \; x > 0$;	$fQ(u) = \frac{1}{\beta}(1 - u)\left\{\log \frac{1}{1-u}\right\}^{1-\beta}$
extreme value	$e^x e^{-e^x}, \; -\infty < x < \infty$;	$fQ(u) = (1 - u)\log \frac{1}{1-u}$
Normal	$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$; $\Phi(x) = \int_{-\infty}^{\infty} \phi(y)\,dy$	$fQ(u) = \frac{1}{\sqrt{2\pi}} \exp -\frac{1}{2}\left|\Phi^{-1}(u)\right|^2 \sim (1 - u)(2 \log \frac{1}{1-u})^{\frac{1}{2}}$

It should be noted that in the β parametrization of exponential type distributions (those for which $\alpha = 1$) the values $\beta = 0$, $.5$, and 1 correspond to the extreme-value, normal and exponential distributions respectively. It should also be noted that the β parametrization does not cover all exponential type distributions; in particular it does not cover

$$\text{Lognormal} \quad f(x) = \frac{1}{x}\,\phi(\log x)\ ; \quad fQ(u) = \phi\Phi^{-1}(u)\ e^{-\Phi^{-1}(u)} \quad .$$

Examples of long tail distributions are

$$\text{Cauchy} \qquad \frac{1}{\pi}\,\frac{1}{1+x^2}\ ,\ -\infty < x < \infty\ ; \quad fQ(u) = \frac{1}{\pi}\cos^2 \pi(u - \tfrac{1}{2})$$

$$= \frac{1}{\pi}\sin^2 \pi u \sim (1 - u)^2$$

$$\text{Reciprocal of a uniform.} \quad \frac{1}{x^2}\ ,\ x > 1\ ; \quad fQ(u) = (1-u)^2$$

$$\text{Pareto}\ \ \beta > 0\ . \quad \left\{\beta\, x^{1+(1/\beta)}\right\}^{-1}\ ,\ x > 1\ ; \quad fQ(u) = \frac{1}{\beta}\,(1 - u)^{1+\beta}$$

$$\text{Tukey}\ \ \lambda < 1\ . \quad Q(u) = \frac{1}{\lambda}\left(u^{\lambda} - (1-u)^{\lambda}\right)\ ; \quad fQ(u) = \frac{(1-u)^{\alpha}}{1 + u^{-\alpha}(1-u)^{\alpha}}\ ,\ \alpha = 1-\lambda$$

The double-exponential distribution exemplifies another aspect of distributions which can be used to classify them — their differentiability.

$$\text{Double-exponential} \quad \frac{1}{2}\,e^{-|x|}; \quad fQ(u) = u \quad \text{for } u < 0.5$$

$$= 1-u \quad \text{for } u > 0.5$$

The non-differentiability (at $x = 0$) of the double exponential density makes the density-quantile function non-differentiable at $u = 0.5$. Non-differentiability of the density is equivalent to the characteristic function

$$\phi(u) = \int_{-\infty}^{\infty} e^{iux}\, f(x)\, dx$$

decaying as $1/u^2$ as $u \to \infty$. Thus one can classify distributions according to the decay rate of (1) their densities and (2) their characteristic functions. The approach to statistical data analysis discussed in this paper basically assumes that the densities we are considering are differentiable in order to obtain reasonable rates of consistency for our estimators.

Given data, the parameters we desire to estimate for it are: location μ , scale σ , tail-exponent α , and (when $\alpha = 1$) shape β .

To efficiently estimate location and scale, one must know $f_0Q_0(u)$ or at least its tail exponent α . A formula given by Andrews (1973) for the tail area of a distribution suggests a fundamental formula for the limiting behavior of fQ functions as $u \to 1$, and also suggests a formula which might be used to rigorously define the tail exponent α of a distribution.

Andrews' tail area approximation formula may be written

$$1 - F(x) = \frac{f(x)}{g(x)} \frac{1}{\kappa - 1} \left[1 + \frac{1}{2}\left\{\frac{g'(x)}{g(x)} - \kappa\right\}\right]$$

defining $g(x) = f'(x)/f(x) = \{\log f(x)\}'$ and

$$\kappa = \lim_{x \to \infty} \frac{g'(x)}{g^2(x)}$$

In this formula, let $u = F(x)$. Then $gQ(u) = -J(u)$, $g'Q(u) = (fQ)''(u)\, fQ(u)$, and

$$1 - u = \alpha \frac{fQ(u)}{J(u)} \left[1 + \frac{1}{2}\left\{\frac{fQ(u)\,(fQ)''(u)}{J^2(u)} - \kappa\right\}\right]$$

defining

$$\alpha = \frac{1}{1-\kappa}, \quad \kappa = \lim_{u \to 1} \frac{fQ(u)(fQ)''(u)}{J^2(u)}$$

The ranges $\alpha < 1$, $\alpha = 1$, $\alpha > 1$ correspond to $\kappa < 0$, $\kappa = 0$, and $\kappa > 0$ respectively.

We are thus led to a <u>rigorous definition of the tail exponent</u> α :

$$\alpha = \lim_{u \to 1} \frac{(1-u)\, J(u)}{fQ(u)}$$

This value of α satisfies approximately for u near 1

$$-\left(\log fQ(u)\right)' = \frac{J(u)}{fQ(u)} = \frac{\alpha}{1-u}$$

whence $\log fQ(u) = \alpha \log (1-u) +$ constant, and

$$fQ(u) \sim (1-u)^{\alpha} \quad ,$$

which is our intuitive definition of α .

One can state a general assumption describing the densities for which the foregoing relations hold. We consider densities $f(x)$ which may have several modes (called multi-modal) but they do not have an infinite number of modes. We call such densities finitely-modal, defined as follows.

A density f is called <u>finitely-modal</u> if: (i) it is non-decreasing on an interval to the right of $a = \sup\{x : F(x) = 0\}$, and it is non-increasing on an interval to the left of $b = \inf\{x : F(x) = 1\}$, where $-\infty \le a < b \le \infty$, and (ii) there is a $\gamma > 0$ such that

$$\sup_{a<x<b} F(x)\left(1 - F(x)\right) \frac{|f'(x)|}{f^2(x)} \le \gamma$$

or equivalently

$$\sup_{0<u<1} u(1 - u) \frac{|J(u)|}{fQ(u)} \le \gamma$$

Finitely-modal densities are considered (without being so named) by Csorgo and Revesz (1978) who demonstrate that they enjoy strong approximations of the quantile process; in Section 10 we apply this fact to estimation of location and scale parameters.

An example of a distribution function which is not finitely-modal is

$$1 - F(x) = \exp(- x - \frac{1}{2} \operatorname{Sin} x)$$

Letting $x = Q(u)$ one obtains a relation for $Q(u)$:

$$-\log(1 - u) = Q(u) + \frac{1}{2} \operatorname{Sin} Q(u)$$

whence $$\frac{1}{1 - u} = q(u) \{1 + \frac{1}{2} \operatorname{Cos} Q(u)\}$$

and $$fQ(u) = (1 - u) \{1 + \frac{1}{2} \operatorname{Cos} Q(u)\}$$

As $u \to 1$, $Q(u) \to \infty$, and $fQ(u)$ oscillates. The hazard quantile function

$$hQ(u) = \frac{fQ(u)}{1 - u} = 1 + \frac{1}{2} \operatorname{Cos} Q(u)$$

also oscillates.

10. Estimation of Location and Scale Parameters

The problem of estimation of location and scale parameters μ and σ usually arises when one assumes that the true distribution function F of X may be represented

$$F(x) = F_0\left(\frac{x-\mu}{\sigma}\right)$$

where F_0 is a known distribution function; we call this representation hypothesis H_0. An equivalent representation may be given for quantile functions:

$$Q(u) = \mu + \sigma Q_0(u) .$$

When F_0 is not known it may be "estimated" from the data using a Goodness of Fit Test for the Hypothesis H_0. We have indicated how to find such goodness of fit tests as a special case of the problem of finding a function Ψ_1, such that $X \sim \Psi_1(Y)$ where Y has a specified distribution F_0. However our approach finds only the derivative $\psi_1(y) = \Psi_1'(y)$ and thus yields only a representation

$$\Psi_1(y) = \mu + \sigma\Psi_0(y)$$

where Ψ_0 is an indefinite integral of ψ_1. Then the quantile function Q of X has the representation

$$Q(u) = \mu + \sigma\Psi_0 Q_0(u)$$

The parameters μ and σ in this representation would be estimated in the same way one estimates any other pair of location and scale parameters.

Much work in the last twenty years has gone into showing how to obtain computationally simple asymptotically efficient estimators of location and scale parameters μ and σ using linear combinations of order statistics. I believe the basic conclusions of this vast effort can be compactly (and even rigorously) summarized by applying the theory of regression analysis on continuous parameter time series from the RKHS (reproducing kernel Hilbert space) point of view given by Parzen (1961), (1967)

A rigorous starting point are the important theorems by Csorgo and Revesz (1978) on strong approximation of the quantile process.

<u>Theorem</u>. Let $X_1,\ldots,X_n$ be i.i.d. random variables with continuous d.f. F and <u>differentiable</u> density f which is finitely-modal and has tail exponent α (as defined at the end of Section 9). The quantile process $\tilde{Q}(u)$ is defined in terms of the order statistics $X_{(1)} < \ldots < X_{(n)}$, and let $\tilde{Q}_U(u)$ be the quantile process of the uniformly distributed random variables $U_j = F(X_j)$. Let

$$R_n = \sup_{0<u<1} \sqrt{n}\ \left| fQ(u)\{\tilde{Q}(u) - Q(u)\} - \{\tilde{Q}_U(u) - u\}\right|$$

Then almost surely

$$\begin{aligned} R_n &= O(n^{-\frac{1}{2}} \log\log n) && \text{if } \alpha < 1 \\ &= O(n^{-\frac{1}{2}}(\log\log n)^2) && \text{if } \alpha = 1 \\ &= O\left(n^{-\frac{1}{2}}(\log\log n)^{\alpha}(\log n)^{(1+\epsilon)(\alpha-1)}\right) && \text{if } \alpha > 1 \end{aligned}$$

where $\varepsilon > 0$ is arbitrary.

To state a theorem concerning the behavior of the uniform quantile process $\tilde{Q}_U$, recall the definition of a Brownian Bridge $\{B(u), \; 0 \le u \le 1\}$; it is a zero mean normal process with covariance kernel

$$K_B(u_1,u_2) = \min (u_1,u_2) - u_1u_2 \quad .$$

Theorem. Csorgo and Revesz (1975). One can define a Brownian Bridge $\{B_n(u), \quad 0 \le u \le 1\}$ for each n such that almost surely

$$\sup_{0\le u\le 1} \left| \sqrt{n} \, \{\tilde{Q}_U(u) - u\} - B_n(u) \right| = O(n^{-\frac{1}{2}} \log n) \quad .$$

For purposes of statistical inference, we can interpret the foregoing results as follows: $\sqrt{n}\; fQ(u) \; \{\tilde{Q}(u) - Q(u)\}$ is distributed as a Brownian Bridge $B(u)$. Under the representation $Q(u) = \mu + \sigma Q_0(u)$ we obtain

$$\sqrt{n} \; \frac{1}{\sigma} f_0Q_0(u) \; \{\tilde{Q}(u) - \mu - \sigma Q_0(u)\} \sim B(u) \quad .$$

Estimating μ and σ becomes a problem in regression analysis of continuous parameter time series by writing

$$f_0Q_0(u) \; \tilde{Q}(u) = \mu \, f_0Q_0(u) + \sigma \, f_0Q_0(u) \; Q_0(u) + \sigma_B \, B(u)$$

where

$$\sigma_B = \frac{1}{\sqrt{n}} \; \sigma \; .$$

We will consider estimators for μ and σ, treating σ_B as a free parameter not constrained to be related to σ. We will find that estimators of σ_B can be used to test the goodness of fit of the model.

The remainder of this section is devoted to writing explicit formulas for asymptotically efficient and unbiased estimators $\hat{\mu}$ and $\hat{\sigma}$, which are linear combinations of order statistics. These formulas assume f_0Q_0 and Q_0 are known; an open problem for research is the use of these formulas with smooth estimators $\widehat{f_0Q_0}$ and $\hat{Q}_0$ to provide adaptive estimators of μ and σ.

Estimating μ and σ given a possibly censored set of order statistics $X_{(np)},\dots,X_{(nq)}$ is more conveniently formulated as using the sample quantile function $\tilde{Q}(u)$ over a subinterval $p \le u \le q$ of $0 \le u \le 1$ (however we permit $p = 0$ or $q = 1$ as possible cases). To form the estimators $\hat{\mu}_{p,q}$ and $\hat{\sigma}_{p,q}$ based on this data we need compute the reproducing kernel inner product $\langle f,g\rangle_{p,q}$ of functions on the interval $p \le u \le q$ corresponding to the kernel $K_B(u_1,u_2)$. We claim that this RKHS consists of L_2 differentiable functions with inner product

$$\langle f,g\rangle_{p,q} = \int_p^q f'(u)\, g'(u)\, du + \frac{1}{p} f(p)\, g(p) + \frac{1}{1-q} f(q)\, g(q) .$$

To verify this assertion, one need only verify the reproducing formula

$$\langle f, K_B(\cdot,t)\rangle_{p,q} = f(t), \quad p \le t \le q .$$

Now

$$\begin{aligned} K_B(u,t) &= u(1-t), & p \le u \le t \\ &= t(1-u), & t \le u \le q \end{aligned}$$

$$\begin{aligned} \frac{\partial}{\partial u} K_B(u,t) &= 1-t, & p \le u \le t \\ &= -t, & t \le u \le q \end{aligned}$$

$$\begin{aligned} \int_p^q f'(u) \frac{\partial}{\partial u} K_B(u,t)dt &= \int_p^t f'(u)(1-t)du + \int_t^q f'(u)(-t)du \\ &= \{f(t) - f(p)\}(1-t) + \{f(q) - f(t)\}(-t) \\ &= f(t) - (1-t)f(p) - tf(q) . \end{aligned}$$

Since $f(p)K_B(p,t) = f(p)p(1-t)$, and $f(q)\,K_B(q,t) = f(q)\,t(1-q)$, we have verified the reproducing property of our formula for inner product.

Define the information matrix

$$I(p,q) = \begin{bmatrix} I_{\mu\mu}(p,q) & I_{\mu\sigma}(p,q) \\ I_{\sigma\mu}(p,q) & I_{\sigma\sigma}(p,q) \end{bmatrix}$$

where

$$I_{\mu\mu}(p,q) = \langle f_0 Q_0, f_0 Q_0 \rangle_{p,q}$$

$$I_{\mu\sigma}(p,q) = I_{\sigma\mu}(p,q) = \langle f_0 Q_0, Q_0(f_0 Q_0) \rangle_{p,q}$$

$$I_{\sigma\sigma}(p,q) = \langle Q_0(f_0 Q_0), Q_0(f_0 Q_0) \rangle_{p,q} .$$

Define

$$T_{n,\mu,p,q} = \langle f_0 Q_0, \tilde{Q}(f_0 Q_0) \rangle_{p,q}$$

$$T_{n,\sigma,p,q} = \langle Q_0(f_0 Q_0), \tilde{Q}(f_0 Q_0) \rangle_{p,q} .$$

Then the optimal estimators are given by

$$\begin{bmatrix} \hat{\mu}_{p,q} \\ \\ \hat{\sigma}_{p,q} \end{bmatrix} = I^{-1}(p,q) \begin{bmatrix} T_{n,\mu,p,q} \\ \\ T_{n,\sigma,p,q} \end{bmatrix}$$

with variance and covariance matrix

$$\begin{bmatrix} \mathrm{Var}(\hat{\mu}_{p,q}) & \mathrm{Cov}(\hat{\mu}_{p,q}, \hat{\sigma}_{p,q}) \\ \\ \mathrm{Cov}(\hat{\mu}_{p,q}, \hat{\sigma}_{p,q}) & \mathrm{Var}(\hat{\sigma}_{p,q}) \end{bmatrix} = \sigma_B^2 \, I^{-1}(p,q) .$$

These estimators may be justified also by their similarity to those given by Weiss and Wolfowitz (1970).

Finally to estimate σ_B^2 we would use an estimator denoted by $\hat{\sigma}^2_{B,p,q}$ which is formed from the residuals $\tilde{Q}(u) - \hat{Q}(u)$; define

$$\hat{Q}_{p,q}(u) = \hat{\mu}_{p,q} + \hat{\sigma}_{p,q} Q_0(u)$$

$$\eta^2_{B,p,q} = \|f_0Q_0(u)\{\tilde{Q}(u) - \hat{Q}(u)\}\|^2_{p,q} .$$

$$\hat{\sigma}^2_{B,p,q} = \frac{1}{n(q-p)} \eta^2_{B,p,q}$$

If we are willing to accept the model, we could take as our estimator

$$\hat{\sigma}^2_{B,p,q} = \frac{1}{n} \hat{\sigma}^2_{p,q}$$

since $\sigma_B^2 = \frac{1}{n} \sigma^2$.

In order to explicitly evaluate the inner product $\langle f,g\rangle_{p,q}$, it is often convenient to use no derivatives of g if one is willing to use second derivatives of f. Since $f'g' + f''g = (f'g)'$ we can write

$$\int_p^q f'g'\,du = -\int_p^q f''g\,du + f'g\Big|_p^q$$

so that

$$\langle f,g\rangle_{p,q} = -\int_p^q f''(u)\,g(u)\,du$$

$$+ g(p)\,[\tfrac{1}{p}\,f(p) - f'(p)]$$

$$+ g(q)\,[\tfrac{1}{1-q}\,f(q) + f'(q)] \;.$$

Thus

$$\langle f_0Q_0,\ \tilde{Q}(f_0Q_0)\rangle = \int_p^q -\{f_0Q_0(u)\}''f_0Q_0(u)\,\tilde{Q}(u)\,du$$

$$+ \tilde{Q}(p)\ f_0Q_0(p)\,[\tfrac{1}{p}f_0Q_0(p) - (f_0Q_0)'(p)]$$

$$+ \tilde{Q}(q)\ f_0Q_0(q)\,[\tfrac{1}{1-q}f_0Q_0(q) + (f_0Q_0)'(q)]$$

$$\langle Q_0(f_0Q_0),\ \tilde{Q}(f_0Q_0)\rangle = \int_p^q -\{Q_0(u)\ f_0Q_0(u)\}''\ f_0Q_0(u)\ \tilde{Q}(u)\,du$$

$$+ \tilde{Q}(p)\ f_0Q_0(p)\ [\tfrac{1}{p}\ Q_0(p)f_0Q_0(p) - \{Q_0(f_0Q_0)\}'(p)]$$

$$+ \tilde{Q}(q)\ f_0Q_0(q)\ [\tfrac{1}{1-q}\,Q_0(q)\ f_0Q_0(q) + \{Q_0(f_0Q_0)\}'(q)] \,.$$

To comprehend the linear functionals in $\tilde{Q}$ which appear in our formulas for $\hat{\mu}$ and $\hat{\sigma}$, define the weight functions

$$W_\mu(u) = -\{f_0Q_0(u)\}'' f_0Q_0(u) = J_0'(u)\, f_0Q_0(u)$$

$$\begin{aligned} W_\sigma(u) &= -\{Q_0(u)\, f_0Q_0(u)\}'' f_0Q_0(u) \\ &= J_0(u) + Q_0(u)\, J_0'(u)\, f_0Q_0(u) \\ &= J_0(u) + Q_0(u)\, W_\mu(u)\,. \end{aligned}$$

Define the additional weights factors

$$W_{\mu L}(p) = f_0Q_0(p)\,[\tfrac{1}{p}\, f_0Q_0(p) + J_0(p)]$$

$$W_{\mu R}(q) = f_0Q_0(q)\,[\tfrac{1}{1-q}\, f_0Q_0(q) - J_0(q)]$$

$$\begin{aligned} W_{\sigma L}(p) &= f_0Q_0(p)\,[\tfrac{1}{p}\, Q_0(p)\, f_0Q_0(p) + Q_0(p)J_0(p) - 1] \\ &= Q_0(p)W_{\mu L}(p) - f_0Q_0(p) \end{aligned}$$

$$\begin{aligned} W_{\sigma R}(q) &= f_0Q_0(q)\,[\tfrac{1}{1-q}\, Q_0(q)\, f_0Q_0(q) + 1 - Q_0(q)\, J_0(q)] \\ &= Q_0(q)W_{\mu R}(q) + f_0Q_0(q)\,. \end{aligned}$$

The linear functionals of $\tilde{Q}$ which appear in $\hat{\mu}$ and $\hat{\sigma}$ may be written

$$T_{n,\mu,p,q} = \int_q^p W_\mu(u)\, \tilde{Q}(u)\, du + \tilde{Q}(p)\, W_{\mu L}(p) + \tilde{Q}(q)\, W_{\mu R}(q)$$

$$T_{n,\sigma,p,q} = \int_p^q W_\sigma(u)\, \tilde{Q}(u)\, du + \tilde{Q}(p)\, W_{\sigma L}(p) + \tilde{Q}(q)\, W_{\sigma R}(q)\,.$$

These integrals are really linear combinations of order statistics if we take $\tilde{Q}(u)$ to be a piecewise constant function equal to $X_{(j)}$ for $(j-1)/n < u \leq j/n$.

The entries of the information matrix may be written:

$$I_{\mu,\mu}(p,q) = \int_p^q |J_0(u)|^2\,du + \frac{1}{p}|f_0Q_0(p)|^2 + \frac{1}{1-q}|f_0Q_0(q)|^2$$

$$= \int_p^q W_\mu(u)\,du + W_{\mu L}(p) + W_{\mu R}(q) .$$

$$I_{\mu,\sigma}(p,q) = \int_p^q J_0(u)\,[Q_0(u)\,J_0(u) - 1]\,du$$

$$+ \frac{1}{p}\,Q_0(p)\,|f_0Q_0(p)|^2$$

$$+ \frac{1}{1-q}\,Q_0(q)\,|f_0Q_0(q)|^2$$

$$= \int_p^q W_\mu(u)\,Q_0(u)\,du + Q_0(p)\,W_{\mu L}(p) + Q_0(q)W_{\mu R}(q)$$

$$= \int_p^q W_\sigma(u)\,du + W_{\sigma L}(p) + W_{\sigma R}(q) .$$

$$I_{\sigma,\sigma}(p,q) = \int_p^q |Q_0 J_0(u) - 1|^2 \, du$$

$$+ \frac{1}{p} |Q_0(p) \; f_0 Q_0(p)|^2$$

$$+ \frac{1}{1-q} |Q_0(q) \; f_0 Q_0(q)|^2$$

$$= \int_p^q W_\sigma(u) \, Q_0(u) \, du + Q_0(p) \, W_{\sigma L}(p) + Q_0(q) \, W_{\sigma R}(q) \; .$$

In the case of a symmetric density $f_0(x) = f_0(-x)$, we have

$$f_0 Q_0(1-u) = f_0 Q_0(u)$$

$$J_0(1-u) = -J_0(u)$$

$$Q_0(1-u) = -Q_0(u) \; .$$

For the case of censorship which is symmetric in the sense that $q = 1-p$, $I_{\mu\sigma}(p,q) = 0$ and

$$\hat{\mu}_{p,q} = \frac{T_{n,\mu,p,q}}{I_{\mu\mu}(p,q)}$$

$$\hat{\sigma}_{p,q} = \frac{T_{n,\sigma,p,q}}{I_{\sigma\sigma}(p,q)} \; .$$

Symmetrically censored normal samples is an important case which we discuss in detail. For the normal distribution

$$f_0Q_0(u) = \frac{1}{\sqrt{2\pi}} \exp -\frac{1}{2}|\Phi^{-1}(u)|^2 ,$$

$$J_0(u) = \Phi^{-1}(u)$$

$$W_\mu(u) = 1$$

$$W_\sigma(u) = 2\Phi^{-1}(u)$$

$$W_{\mu R}(q) = f_0Q_0(q)[\frac{1}{1-q} f_0Q_0(q) - \Phi^{-1}(q)]$$

$$W_{\sigma R}(q) = f_0Q_0(q)[1 + \Phi^{-1}(q)\{\frac{1}{1-q} f_0Q_0(q) - \Phi^{-1}(q)\}] .$$

In order to study the behavior of these weights as $q \to 1$, we note an important property of the normal distribution [which follows from Feller, Vol. 1, p. 166, eq. (1.8)]:

$$0 \leq \frac{1}{1-q} f_0Q_0(q) - \Phi^{-1}(q) \leq \frac{1}{1-q} f_0Q_0(q)\{\Phi^{-1}(q)\}^{-2}$$

which tends to 0 as q tends to 1.

11. Some open research problems for extensions

The approach described in this paper can be described as one which formulates statistical estimation and testing problems as problems of density estimation and testing for white noise. This paper discussed only the univariate one-sample case. Two-sample and multivariate (including non-parametric regression) problems can be treated similarly (see Parzen (1977)). This section describes some extensions of our results in the one-sample case whose theory and application is open for research.

Power Transformation to Normality. The transformation of a random variable X to a $N(\mu,\sigma^2)$ distribution is often assumed to be of the form

$$\Psi(x) = \frac{1}{\lambda}\{(x-\zeta)^{\lambda} - 1\}, \qquad \lambda \neq 0$$

$$= \log(x-\zeta), \qquad \lambda = 0 .$$

The derivative $\psi(x) = \Psi'(x)$ has a single formula

$$\psi(x) = (x-\zeta)^{\lambda-1} .$$

The quantile function $Q(u)$ of X is then related to the standard normal quantile function $\Phi^{-1}(u)$ by

$$\mu + \sigma\Phi^{-1}(u) = \frac{1}{\lambda}\{(Q(u)-\zeta)^{\lambda} - 1\}, \qquad \lambda \neq 0$$

$$= \log(Q(u)-\zeta), \qquad \lambda = 0 .$$

The density-quantile function of X satisfies

$$\log fQ(u) = -\log \sigma + \log \phi\Phi^{-1}(u) + (\lambda - 1) \log (Q(u) - \zeta)$$

The problem is: (1) to use these relations to estimate the parameters λ and ζ ; and (2) compare these estimators with the estimators of Box and Cox (1964).

Survival data. Let $X_1,\ldots,X_n$ be a random sample from a single lifetime or survival distribution F with quantile function Q . However one may fail to observe an X (called a "death") due to the previous occurrence of some other event Y (called a "loss") which has distribution H . The desired value X is censored on the right by Y , and one observes

$$Z = \min (X,Y)$$

with distribution function G satisfying

$$1 - G = (1 - F)(1 - H)$$

under suitable independence assumptions.

From the observed data $Z_1,\ldots,Z_n$ one can form an estimator $\tilde{F}$ of F introduced by Kaplan and Meier (1958). Its quantile function $\tilde{Q}$ is an estimator of Q . The asymptotic distribution theory of $\tilde{F}$ and $\tilde{Q}$ has been found by Breslow and Crowley (1974) and Sanders (1975) respectively; the latter shows that $\sqrt{n}\, fQ(u) \{Q(u) - \tilde{Q}(u)\}$, $0 < u < 1$, converges in distribution (as a stochastic process) to a zero mean Gaussian process with covariance kernel K given by

$$K(u_1,u_2) = (1 - u_1)(1 - u_2) \int_0^{\min(u_1,u_2)} dw\ (1 - w)^{-2} \{1 - HQ(w)\}^{-1}$$

When there is no censoring, $H = 0$ and $K(u_1,u_2) = u_1(1 - u_2)$ for $u_1 < u_2$, the covariance kernel of the Brownian bridge.

The covariance kernel K has an integral representation which makes it easy to find its RKHS inner product. Thus one would have no difficulty extending the results of Section 10 to estimation of location and scale parameters from survival data.

Sampling the Quantile Process. Suppose that to compress the data one seeks to reduce a sample of size n to k values, namely the order statistics $X_{(np_j)} \doteq \tilde{Q}(p_j)$ corresponding to specified percentiles $p_1,\ldots,p_k$. One can choose these percentiles so that the optimal linear estimators $\hat{\mu}$ and $\hat{\sigma}$ that could be formed from them have variances which are a minimum over all choices of k points at which to sample $\tilde{Q}(u)$. Results of this kind could be deduced from the work of Sacks and Ylvisaker (1966) on designs of continuous parameter time series regression problems.

References

Section 1.

Tukey, J. W. (1977) Exploratory Data Analysis, Addison-Wesley: Reading, MA

Section 2.

Elderton, W. P. and Johnson, N. L. (1967) Systems of Frequency Curves, Cambridge University Press

Hajek, J. and Sidak, Z. (1967) Theory of Rank Tests, Academic Press, New York

Roussas, G. C. (1973) A First Course in Mathematical Statistics, Addison-Wesley: Reading, MA.

Section 3.

Barlow, R. E. and Campo R. (1975) Total time on test processes and applications to failure data analysis. Reliability and Fault Tree Analysis, ed. R. E. Barlow, J. Fussell, N. Singparwalla, SIAM : Philadelphia

Barlow, R. E. and Doksum, K. A. (1972) Isotonic tests for convex orderings. Proc. Sixth Berkeley Sym. on Math. Statist. and Prob., Vol. I, 293-323

Barlow, R. E. and Proschan, F. (1977) Asymptotic theory of total time on test processes with applications to life testing. Multivariate Analysis IV, ed. P. Krishnaiah, North Holland : Amsterdam, 227-237

Section 4.

Geronimus, Y. L. (1960) Polynomials Orthogonal on a Circle and Interval (translated from Russian), Pergamon : New York.

Section 5.

Chernoff, H., Gastwirth, J. L. and Johns, M. V. Asymptotic distribution of linear combinations of order statistics with applications to estimation. Ann. Math. Statist., 38, 52-72

Czorgo, M. and Revesz, P. (1975) Some notes on the empirical distribution function and the quantile process. Limit Theorems of Probability, North Holland : Amsterdam, 59-71

Czorgo, M. and Revesz, P. (1978) Strong approximations of the quantile process. Annals of Statistics.

Moore, D. S. (1968) An elementary proof of asymptotic normality of linear functions of order statistics. Ann. Math. Statist., 39, 263-265

Pyke, R. (1965) Spacings (with discussion). J. Roy. Statist. Soc. B, 27, 395-449

Pyke, R. (1972) Spacings Revisited, Proceedings Sixth Berkeley Symposium on Mathematical Statistics and Probability, ed. L. LeCam and J. Neyman, University of California Press, I, 417-427

Shorack, G. R. (1972) Convergence of quantile and spacings processes with applications. Ann. Math. Statist., 43, 1400-1411

Stigler, S. M. (1974) Linear functions of order statistics with smooth weight functions. Ann. Statist., 2, 676-693

Stigler, S. M. (1977) Fractional order statistics, with applications. Jour. Amer. Statist. Assn., 544-556

Section 6.

Barlow, R. E. and Van Zwet, W. R. (1970) Asymptotic Properties of Isotonic Estimators for the Generalized Failure Rate Function. Part I: Strong Consistency, in M. L. Puri, ed., Nonparametric Techniques in Statistical Inference, Cambridge Univ. Press, 159-173

Davis, H. T. and Jones, R. H. (1968) "Estimation of the innovation variance of a stationary time series." Journal of the American Statistical Association, 63, 141-149

Durbin, J. (1973) Distribution Theory for Tests Based on the Sample Distribution Function, Regional Conference Series in Applied Mathematics, 9, SIAM : Philadelphia

Fisher, R. A. (1929) "Tests of significance in harmonic analysis," Proc. Roy. Soc. London, Ser. A, 125, 54-59

Gnanadesikan, R. (1977) Methods for Statistical Analysis of Multivariate Observations, Wiley : New York

Hannan, E. J. and Nicholls, D. F. (1977) The estimation of the prediction error variance, Journal of the American Statistical Association, 72, 834-840

Mann, N. R. and Fertig, K. W. (1975) A goodness-of-fit test for the two parameter vs. three parameter Weibull; confidence bounds for threshold. Technometrics, 17, 237-245

Weiss, L. (1963) On the asymptotic distribution of an estimate of a scale parameter. Naval Research Logistics Quarterly, 10, 1-9

Section 7.

Bofinger, E. (1975) Estimation of a density function using order statistics, Austral. J. Statist., 17, 1-7

Carmichael, J. P. (1976) The Autoregressive Method, Ph.D. Thesis, Statistical Science Division, State University of New York at Buffalo

Moore, D. S. and Yackel, J. W. (1977) Consistency properties of nearest neighbor density function estimators. Ann. Statist., 5, 143-154

Parzen, E. (1967) Time Series Analysis Papers, Holden Day : San Francisco.

Parzen, E. (1974) Some Recent Advances in Time Series Modeling, IEEE Transactions on Automatic Control, AC-19, 723-730

Parzen, E. (1977) Multiple Time Series: Determining the Order of Approximating Autoregressive Schemes, Multivariate Analysis - IV, edited by P. Krishnaiah, North Holland: Amsterdam, 283-296

Wahba, G. and Wold, S. (1975) Periodic splines for spectral density estimation: the use of cross-validation for determining the correct degree of smoothing. Commun. Statist., 4, 125-141

Section 8.

Thaler, H. (1974) Non-parametric probability density estimation and the empirical characteristic function, Ph.D. Thesis, Statistics Department, State University of New York at Buffalo

Section 9.

Andrews, D. F. (1973) A general method for the approximation of tail areas. Annals Statistics, 1, 367-372

Gumbel, E. J. (1962) Statistical theory of extreme values. Contributions to Order Statistics, ed. Sarhan, A. E. and Greenberg, R. E., Wiley: New York, 56-93

Sarhan, A. E. and Greenberg, B. G. (1962) Contributions to Order Statistics, Wiley: New York

Section 10.

Parzen, E. (1961) An approach to time series analysis. Ann. Math. Statist., 32, 951-989

Parzen, E. (1961) Regression analysis of continuous parameter time series, Proc. 4th Berkeley Sympos. Math. Statist. and Prob., Univ. California Press, Berkeley, Calif., Vol. I, 469-489

Parzen, E. (1970) Statistical inference on time series by RKHS methods, II, Proceedings of the 12th Bienniel Canadian Mathematical Congress, edited by R. Pyke, American Mathematical Society, Providence, R.I., 1-37

Weiss, L. and Wolffowitz, J. (1970) Asymptotically Efficient Non-Parametric Estimators of Location and Scale Parameters, Z. Wahrscheinlichkeits-Theorie verw. Gebiete, 16, 134-150

Section 11.

Box, G. E. P. and Cox, D. R. (1964) An analysis of transformations. J. Roy. Stat. Soc., B26, 211-252

Breslow, N. and Crowley, J. (1974) A large sample study of the life table and product limit estimates under random censorship. Ann. Statist., 2, 437-453

Kaplan, E. L. and Meier, P. (1958) Nonparametric estimation from incomplete observations. J. Am. Statist. Assoc., 53, 457-481

Parzen, E. (1977) Nonparametric Statistical Data Science: A Unified Approach Based on Density Estimation and Testing for "White Noise." Technical Report No. 47, Statistical Science Division, SUNY at Buffalo

Sacks, J. and Ylviskar, D. (1966) Designs for regression problems with correlated errors. Ann. Math. Statist., 37, 66-89

Sander, J. M. (1975) The weak convergence of quantiles of the product-limit estimator. Stanford University Technical Report No. 5